UNIVERSITY SYSTEM OF GEORGIA

ARTIFICIAL INTELLIGENCE GUIDELINES: A USG IT HANDBOOK COMPANION GUIDE

VERSION 1.1

6/2/2024

PUBLIC

This companion guide is to aid USG organizations when contemplating, procuring, implementing and maintaining Artificial Intelligence oriented solutions.

Revision & Sign-off

Change Record

Date	Author	Version	Change Reference
20240712	Todd Watson	1.1	Initial release

Document Properties

Item	Details
Document Title	Artificial Intelligence Guidelines: A USG IT Handbook Companion Guide
Document Type	Guideline (Internal Use Only)
Author	Dr. Todd Watson
Document Manager	Dr. Todd Watson
Creation Date	20240602
Last Updated	20240712
Document Classification	Public

Table of Contents

Introduction

Artificial intelligence (AI) technologies evolve and will continue to be integrated into critical business functions. USG organizations must employ AI responsibly and ethically. Recognizing that AI tools are anticipated to be essential for accurate and timely business functions, it is important to identify and mitigate risks or undesirable outcomes during project planning.

This guide is intended to identify considerations for deploying general AI solutions, Generative AI (GenAI), Deep Learning (DL) and Machine Learning (ML) tools within organization research, services and operations.

TERMS AND DEFINITIONS

- **Artificial Intelligence (AI) –** A technology family that enables computers to perform a variety of advanced functions, including the ability to process visual cues, understand and translate spoken and written language, analyze data, and make recommendations from heuristic analyses.
- **Deep learning (DL) –** A method of AI that teaches computers to process data in a way that is inspired by the human brain. Deep learning models typically are used to recognize complex patterns in pictures, text, sounds and other data to produce accurate insights and predictions.
- **Generative AI (GenAI) –** A form of AI capable of generating text, images, videos or other data using generative models, often in response to prompts.
- **Hallucinations -** Conditions when an LLM process identifies patterns or objects that are nonexistent, creating nonsensical or inaccurate outputs.
- **Large Language Model (LLM) –** A computational model recognized for the ability to achieve general-purpose language generation and other natural language processing tasks such as classification.
- **Machine Learning (ML)** - A branch of AI and computer science that focuses on using data and algorithms to enable AI to imitate the way that humans learn, gradually improving its accuracy
- **Prompt Injection** – A specialized type of cyber-attack against LLMs, whereby bad actors disguise malicious inputs as legitimate, resulting in the return of erroneous results or leaking sensitive information.

1. Responsibilities

Higher education is a pioneer for harnessing the potential of AI to enhance services, improve efficiency, and drive innovation. While AI research has been underway since the 1950s, a recently emerging subdiscipline garnering public attention is GenAI. Abiding principles for USG AI adoption include:

1. *Protection*: Protecting and preserving the safety of humans.
2. *Compliance*: Complying with law and similar regulations.
3. *Focus*: Using AI for the planned purpose.
4. *Purposed*: Advancing USG's mission, values and business objectives.
5. *Performance*: Periodically reviewing results with stakeholders to ensure AI is successfully meeting objectives.
6. *Vigilance*: Employing due care when using AI to safeguard against abuse.

7. *Inventoried*: Maintaining a current inventory of all AI components and tools that detail their type, purpose and organization benefit.
8. *Human-Informed*: Because AI tools often mimic human behavior, speech, and mannerisms, AI interacting directly with humans should disclose its outputs or responses are machine-generated. In circumstances where AI is used in research with human subjects, disclosing the use of AI should be incorporated into informed consent.

Suppliers

Suppliers, contractors and subservice providers of AI should provide a written philosophical and operational framework explaining the responsible use of AI (RAI) for review and approval prior to the execution of any services that leverage AI. Agreements for AI services should include supplier's adherence to their RAI principles, including:

1. The stated purpose and operation of the AI solution.
2. Supplier's methods of data acquisition, preparation, enhancement, detection and mitigation of unacceptable AI results.
3. Technologies employed to detect AI inaccuracy, bias and other unacceptable results.
4. Review of software engineering best test practices to ensure the AI solution is functioning as intended and is trustworthy, including:
    a. Conducting rigorous unit tests of components in isolation.
    b. Conducting integration tests to understand how individual AI components interface with USG systems.
    c. Perform statistical testing of inputs to detect drift or unexpected results.
    d. Refine reference test data regularly to reduce the likelihood of the AI model training on the test set.
    e. Conduct iterative user testing to ensure user needs are being achieved in the development cycle.
    f. Employing quality engineering to avoid unexpected surprises. Quality checks should detect unintended failures and trigger an immediate response.
5. Use of a reference "gold standard" data set against the output to ensure predictable outcomes.
6. Model monitoring, including use of multiple metrics to assess training and monitoring.
7. For models that cannot be adequately explained, a detailed description of the intended outcomes to enable a review of the AI or ML model outputs.
8. Means for detecting and mitigating biases:
    a. Description and source of data used to train AI models.
    b. Description of data and classification of data being used.
    c. Description of how bias is detected and mitigated (or adjusted for bias.
    d. Description of elements in creating the model, including parameters and algorithms used with an indication of parameters with the most influence.

Organization

Adoption of AI tools including GenAI, should progress responsibly. Because GenAI can produce unintended consequences, the principles of ethics, accountability, transparency and protection of humans must be contemplated during the design and operational phases. USG organizations are responsible for ensuring that AI technologies avoid harm, respects rights of individuals and adhere to the highest ethical standards, including:

1. Identifying a responsible person with sufficient ongoing authority over each AI tool or service to authorize and deauthorize operation.
2. Developing a plan detailing how AI products or developed solutions will be used to address specific problems.
3. Performing ongoing testing to ensure the safety, accuracy, and reliability of data and algorithms.
4. Performing ongoing reviews to identify ethical, legal, and societal issues arising from the AI service.
5. Developing an ongoing monitoring and maintenance plan to ensure that AI systems operate safely and effectively as the knowledge model evolves. Plans should include processes to identify and address issues and concerns promptly.
6. Procuring AI solutions should in all other manners adhere to USG legal, ethical, business and technology guidelines.
7. Implementing AI solutions should apply the NIST AI Risk Management Framework and USG business and technology standards when assessing potential suppliers for AI-enabled solutions.

Data Protection

To protect data from exposure, USG Organizations should:

1. Ensure AI solutions are approved for use in the manner proposed and that all confidential or sensitive information used in conjunction with AI is protected, encrypted or anonymized.
2. Implement appropriate security measures to protect data from unauthorized access, modification, or deletion.
3. Consider and safeguard AI solutions from deriving confidential information from otherwise disparate data.
4. Report data breaches involving AI according to the organization Incident Response Plan.

External Collaboration

When developing AI tools or engaging with external collaborators, agreements should:

1. Disclose the purpose and intended use of AI in the collaboration.
2. Define roles, responsibilities and expectations of all parties involved.
3. Identify ownership, access and usage rights of AI models and data.
4. Prioritize safety, privacy, and ethical considerations.
5. Undertake effort to identify and mitigate potential biases by auditing and using unbiased data to train the model(s).
6. Commit to transparency, accountability, and respect for privacy.
7. Commit to pre-deployment testing to ensure principles in this guide are adhered to.
8. Be reviewed for potential ethical, legal, and societal implications.

2. Training

System users should be trained beforehand on the responsible and ethical use of AI and ways to prevent intentional and unintentional use and abuse, including:

1. Familiarization with the AI standard in *USG IT Handbook* in addition to relevant policies, guidelines and best practices.

2. Participation in regular training sessions and workshops on the evolving developments of AI and the emerging risks.

3. Bias

To control potential biases in AI-generated outputs, USG organizations should:

1. Be aware of typical forms of bias potential within AI systems, to include training data, cognitive, algorithmic and confirmation biases.
2. Regularly review and evaluate outputs for potential biases and inaccuracies, seeking input from diverse perspectives and stakeholder groups.
3. Review methodologies to better understand the AI tool's processes for creating results.
4. Collaborate with AI suppliers and developers to reduce unacceptable bias.

4. Accuracy

To ensure that AI-generated outputs are accurate and appropriate, USG organizations should:

1. *Check accuracy* - Confirm outputs are accurate by reviewing the outputs with independently verifiable methods.
2. *Review appropriateness* – Confirm outputs are appropriate for the intended purpose.
3. *Review reasonableness* – Review results by the responsible person or a designee prior to publishing or using the results in a substantive decision.
4. *Review compliance* - Develop and implement guidelines for AI-generated outputs according to the contexts and situations of the organization. Determinants of compliance should consider potential risks from hallucinations, prompt injections, copyright infringements and exposing intellectual property.
5. *Perform recordkeeping* - USG organizations should ensure actions taken based on the AI-generated results are memorialized.

5. Misuse

USG organizations should report any data breaches or incidents involving AI systems promptly in accordance with their *Incident Response Plan* and the *USG IT Handbook*. USG organizations should be aware of the potential for misuse and mitigate situations pertaining to intentional and unintentional misuse without unreasonable delay.

Intentional misuse

AI tools can intentionally be misused. USG organizations should be mindful of and guard against intentional misuse, including:

1. *Fraud*: Using AI to manipulate or cheat otherwise unsuspecting individuals or organizations using scams or other fraudulent behavior, including financial and academic fraud.
2. *Invasion of Privacy:* Using AI systems to gather personal information without an individual's permission.
3. *Malicious use:* Leveraging AI systems for cyberattacks, including vulnerability identification/exploitation, phishing attempts or social engineering.

4. *Propagating misinformation:* Leveraging AI systems to mislead or create and distribute false information.
5. *Discrimination*: Using AI systems to create bias or discriminatory situations that result in unequal treatment of individuals or groups.

Unintentional misuse

AI tools can unintentionally be misused in several ways. USG organizations should be mindful of and guard against unintentional misuse of AI, including:

1. *Data Exposure*: Sensitive or personal information from controlled data supplied to the AI system, failure to de-identify or encrypt information properly, or identification of persons or sensitive information derived from non-sensitive data sources.
2. *Misleading, incorrect conclusions or information:* AI-generated outputs that are incorrect, outdated, or misleading, thus contributing to financial losses, reputational damage or poor decisions.
3. *Inappropriate results:* AI systems producing biased, offensive or otherwise inappropriate content for the intended purpose.
4. *Overdependency on AI:* USG organizations may rely on AI systems and fail to apply human judgment, expertise and common sense. Over-reliance on AI risks adopting poor solutions, overlooking human perspectives and producing subpar outcomes.
5. *Bias or discrimination*: AI systems inadvertently producing biases or discriminatory patterns, leading to unfair treatment of individuals or groups.
6. *Drift*: Include a plan for ongoing monitoring and maintenance to ensure that the original purpose of the system does not drift, and identified concerns are mitigated.

6. Termination and Business Continuity

After an appropriate investigation, USG CIO or CISO may require suspension of any AI tool that is confirmed to generate operational, reputational, data privacy, security, or malicious consequences, whereupon the USG organization must cease operation of the AI tool, investigate, and if warranted, develop and execute a mitigation plan addressing the risk prior to resuming operation.

USG Organizations should have a business continuity plan for alternate means for achieving results obtained by critical AI tools.